# Domain Adaptation for Medical Image Segmentation using Transformation-Invariant Self-Training

**Negin Ghamsarian[1], Javier Gamazo-Tejero[1], Pablo Márquez-Neila[1], Sebastian Wolf[2], Martin Zinkernagel[2], Klaus Schoeffmann[3], Raphael Sznitman[1]**

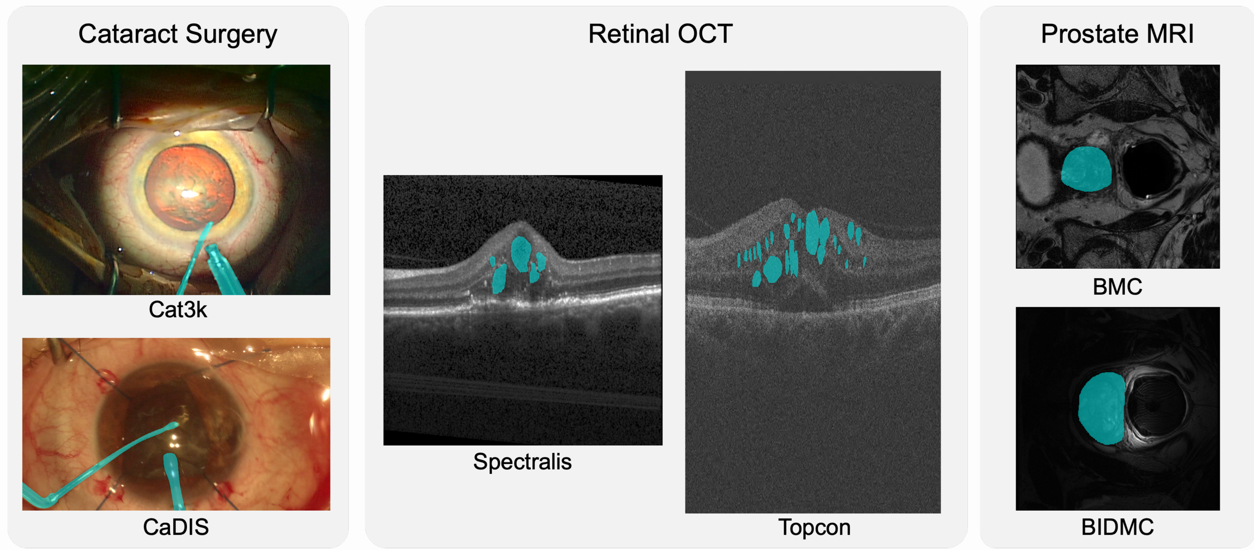
[1]ARTORG Center for Biomedical Engineering Research, Universität Bern, Bern, Switzerland
[2]Department of Ophthalmology, Inselspital, Bern, Switzerland
[3]Institute of Information Technology, Universität Klagenfurt, Klagenfurt, Austria

UNIVERSITÄT BERN

## Abstract

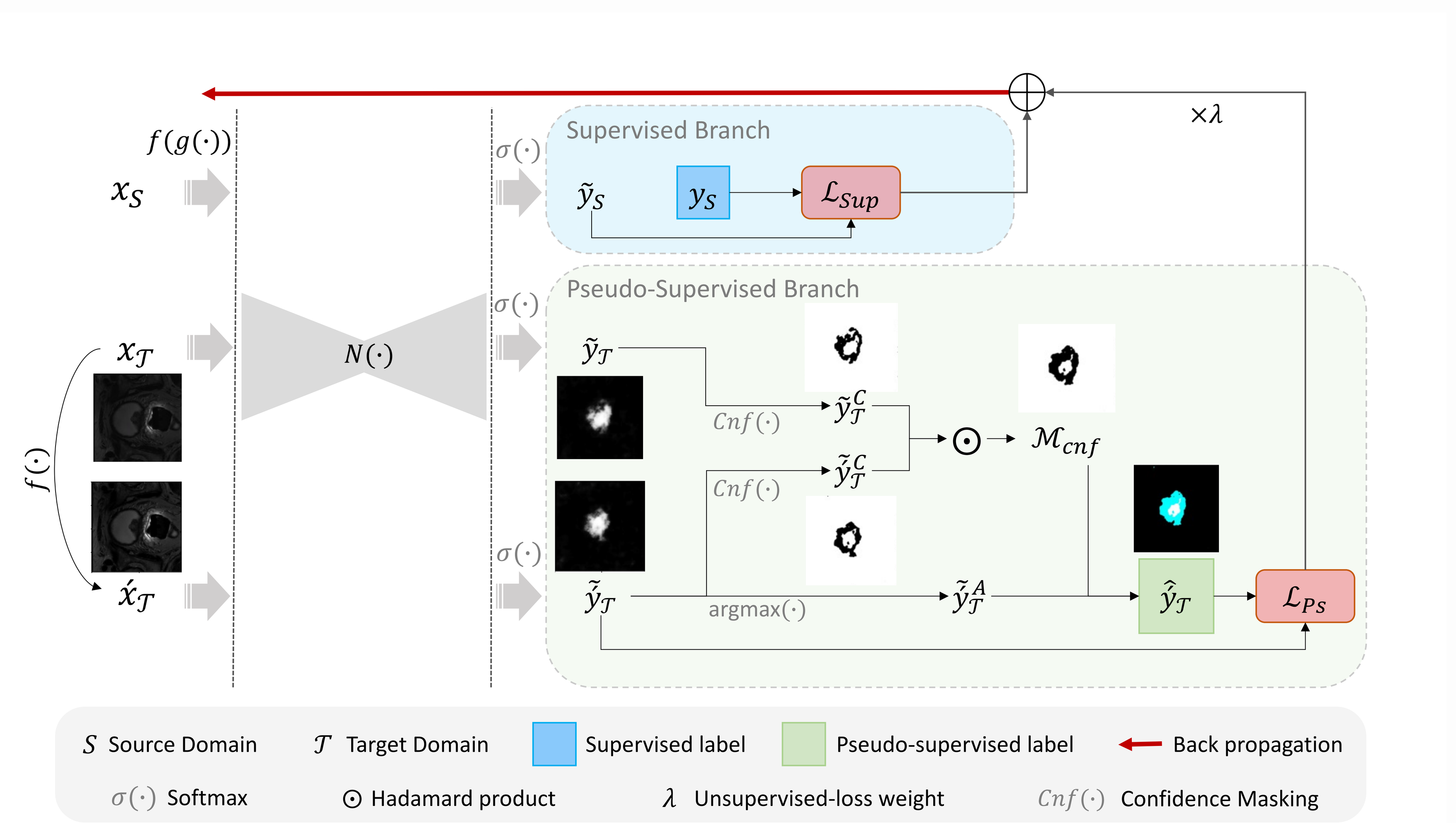

- Leveraging unlabelled data is crucial for addressing distribution gaps in diverse image datasets.
- Self-training techniques using pseudo-labels are highly effective for semi-supervised domain adaptation.
- Unreliable pseudo labels can hinder self-training's performance, especially in the case of significant distribution gaps.
- Identifying uncertain pseudo labels through image transformation variance can improve ground truth approximations.
- The proposed "transformation-invariant self-training (TI-ST)" filters out unreliable pseudo-labels, enhancing domain adaptation for medical image segmentation.


Example images from the three adopted cross-device-and-center datasets:




## Methodology

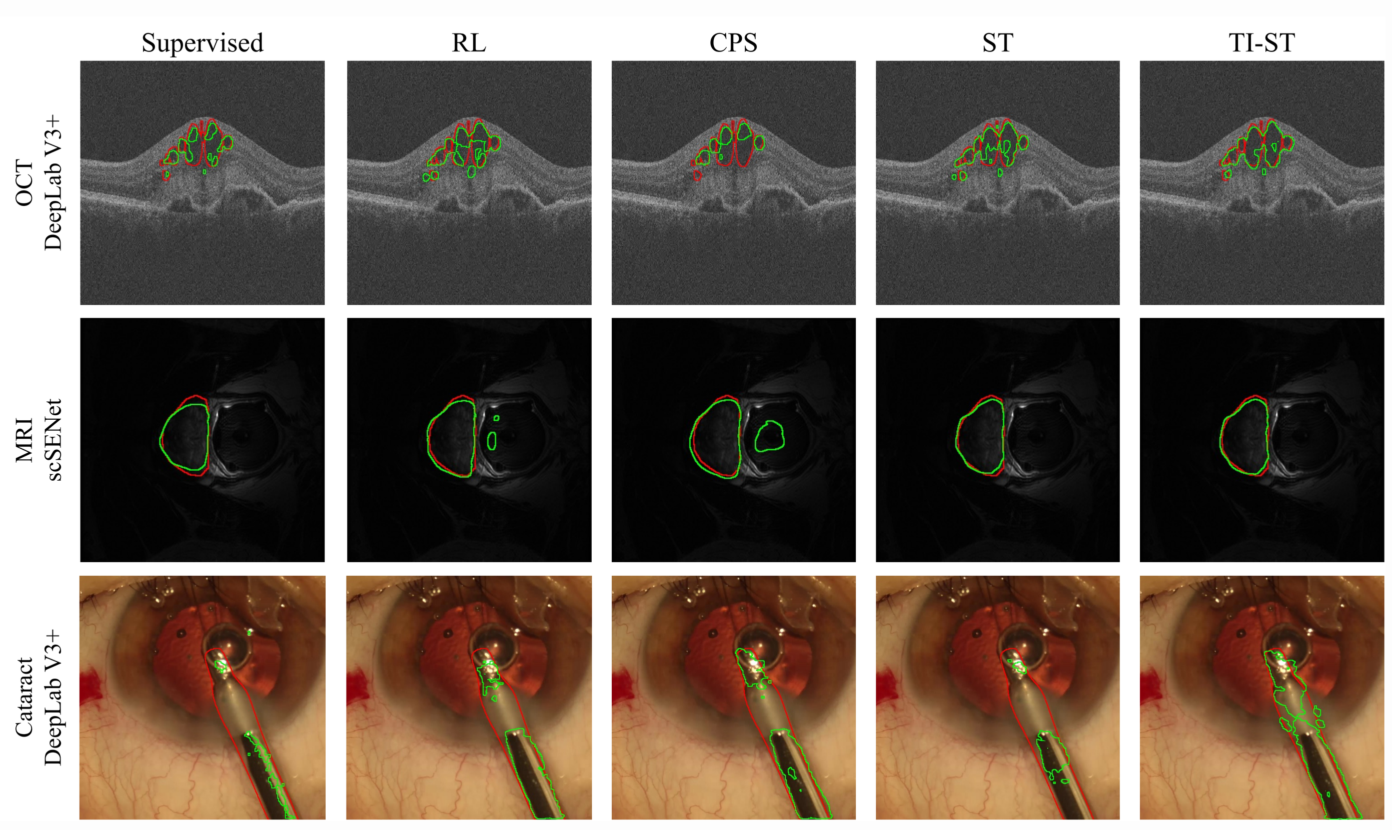
Overview of the proposed semi-supervised domain adaptation framework based on transformation-invariant self-training (Ignored pseudo-labels during unsupervised loss computation are shown in turquoise):




- The goal is to train a model for semantic segmentation in the target dataset using both labeled source and unlabeled target datasets.
- TI-ST assigns pseudo labels to target images during training, but with a self-assessment strategy for reliability estimation.
- TI-ST focuses on retaining high-confidence predictions and filters out transformation-variant predictions.
- The network is simultaneously trained on a batch of labeled and a batch of unlabeled images.
- Images from the target dataset are fed in two versions, the original and non-spatially transformed, and their predictions are computed.
- A confidence-mask ensemble is formed to encode regions of confident predictions that are invariant to transformations.
- The training loss combines supervised and pseudo-supervised losses, with a gradual increase in the weight of pseudo-supervised loss to reinforce training on transformation-invariant highly-confident predictions.

## Qualitative Comparisons

Qualitative comparisons between the performance of TI-ST and the best alternatives:

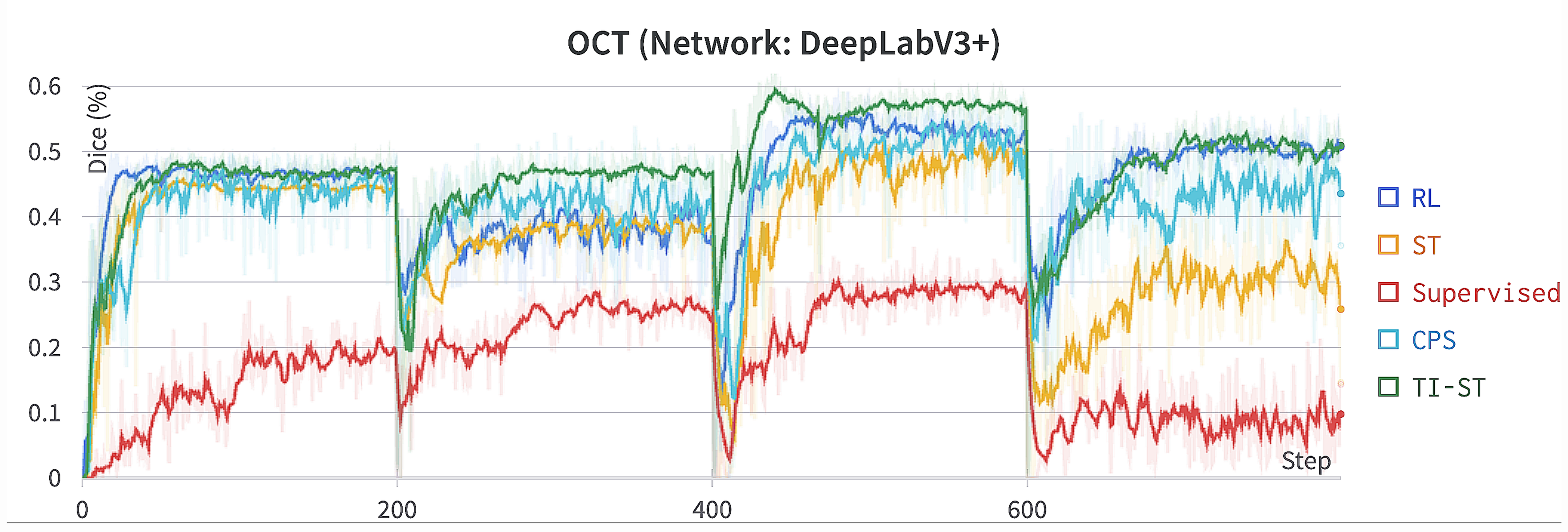



## Quantitative Comparisons

Quantitative comparisons in Dice score (%) among the proposed (TI-ST) and alternative methods for DeepLabV3+ (DLV3+) and scSENet and the three datasets (Relative Dice computed over the Supervised baseline. The best results are shown in green):

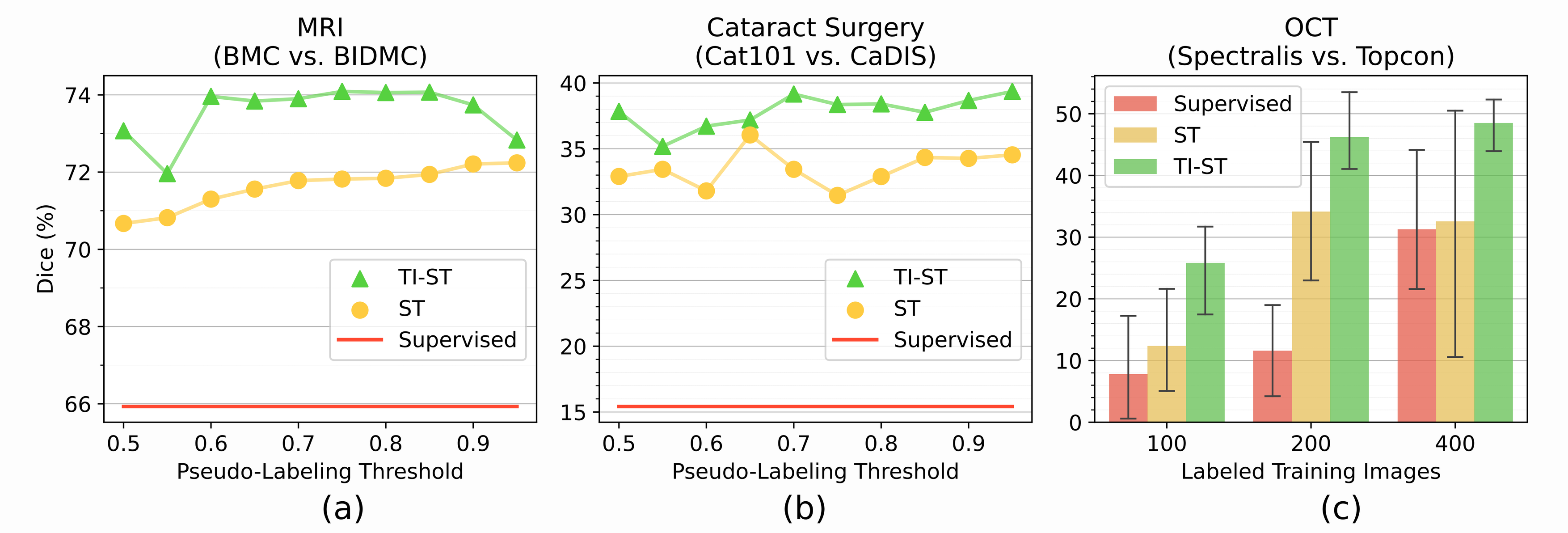
| Modality | Cataract Surgery | | OCT | | MRI | | Avg. Rel. |
|---|---|---|---|---|---|---|---|
| Network | DLV3+ | scSENet | DLV3+ | scSENet | DLV3+ | scSENet | |
| Supervised | 15.42 | 37.67 | 22.87 | 24.08 | 52.39 | 65.93 | N/A |
| Π Model | 27.55 | 35.56 | 1.12 | 0.00 | 10.00 | 6.87 | -22.88 |
| TE | 33.10 | 42.32 | 42.13 | 39.86 | 63.41 | 67.25 | 11.62 |
| Mean Teacher | 11.06 | 39.54 | 19.11 | 4.70 | 64.82 | 66.87 | -2.04 |
| RL | 34.40 | 45.13 | 48.73 | 47.70 | 60.79 | 70.20 | 14.77 |
| CPS | 36.24 | 39.40 | 47.31 | 14.71 | 76.00 | 68.80 | 10.68 |
| ST | 34.34 | 41.10 | 36.84 | 33.01 | 68.63 | 71.97 | 11.26 |
| MCF | 26.97 | 40.19 | 40.12 | 36.52 | 54.17 | 50.23 | 7.46 |
| **TI-ST** | 37.69 | 45.31 | 50.93 | 40.87 | 66.56 | 74.07 | 16.18 |
| | (+22.27) | (+7.46) | (+28.06) | (+16.79) | (+14.17) | (+8.14) | |

Four-fold training curves corresponding to TI-ST and the main alternative methods:

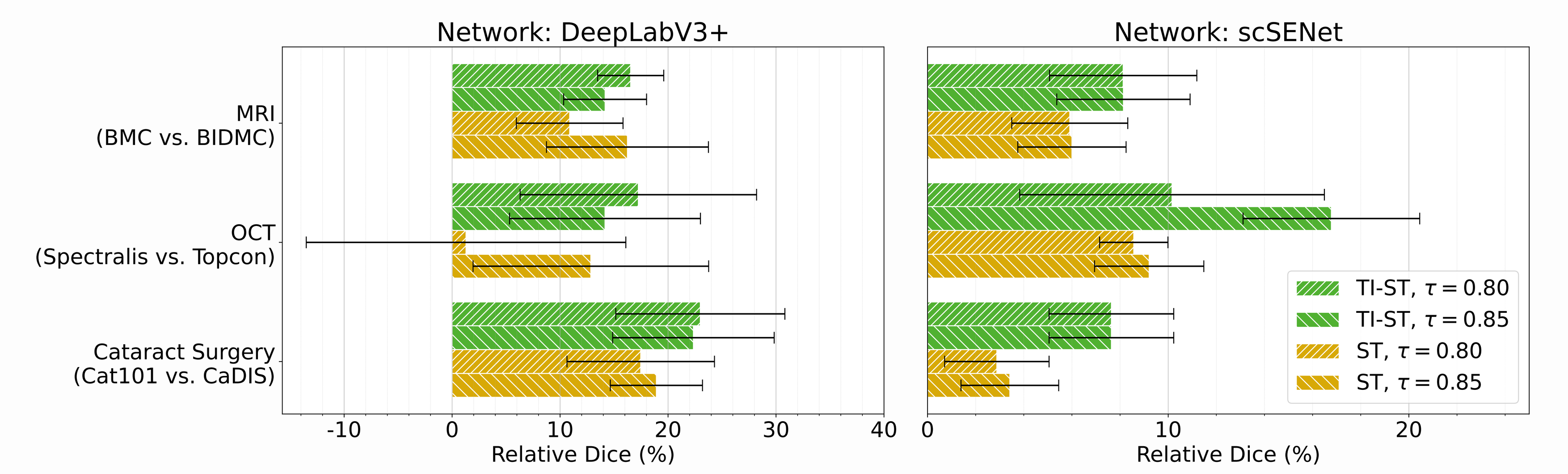



## Ablation Studies

Ablation studies on the pseudo-labeling threshold and size of the labeled dataset:




Ablation study on the performance stability of TI-ST vs. ST across the different experimental segmentation tasks:




MICCAI 2023 Vancouver CANADA

This work was funded by Haag-Streit Foundation, Switzerland.